



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ

ЛАБОРАТОРНА РОБОТА №1

З дисципліни «Криптографія»

Виконали:

студенти 3 курсу ФТІ

групи ФБ-93

Абдуллаєва Есміра

Шовак Мирослав

Викладач:

Селюх П. В.

Мета роботи: засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

Завдання

1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H_1 та H_2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H_1 та H_2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилучено всі пробіли.

2. За допомогою програми CoolPinkProgram оцінити значення $H(10)$, $H(20)$, $H(30)$.

3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Хід роботи та опис труднощів:

Під час роботи, з огляду на вимоги форматування тексту, та різних операцій над ним, було вирішено використовувати мову python. У якості вхідного тексту була обрана книга "Зелена миля". Форматування вхідного тексту не викликало значних труднощів. За допомогою стандартних функцій ми змогли позбутися пробілів, твердого знаку, змінити всі великі букви на малі. За допомогою регулярного виразу змогли досягнути зчитування з файлу тільки алфавіту. Для зручнішого використання програми, в залежності від завдання, було вирішено зробити лаконічне меню з 3 можливостями: зчитування файлу з пробілами і виконання операцій над знаками, зчитування файлу без пробілів, аналогічно було проведено для біграм. Сортування монограм ми досягли за допомогою numpy. Труднощі виникли при створенні функції для зчитування біграм. Було прийнято рішення створити масив букв, і вже до значень цього масиву додавати, як аргумент функції ще одну букву по черзі і потім рахувати утворені біграми у списку, який ми отримали за допомогою `line = [one[k:k + 2] for k in range(0, len(one), 2)]` (випадок з неперехресними біграмами) та `line = [one[k:k+2] for k in range(len(one))]` (випадок з перехресними). Вивід реалізовували у вигляді квадратної матриці, для презентабельного вигляду ми відформатували значення наших H_2 до трьох знаків після коми.

Код програми

```
import math
import re
import numpy as np
file = open("1.txt", "r")
data = file.read().replace("ъ", "ь")
one = re.sub(r"[^а-я]+", " ",
data.lower()).replace("ъ", "ь").replace("ё", "е")

# m=32 - кількість букв алфавіту
alphabet =
['а', 'б', 'в', 'г', 'д', 'е', 'ж', 'з', 'и', 'й', 'к', 'л',
'м', 'н', 'о', 'п', 'р', 'с', 'т', 'у', 'ф', 'х', 'ц', 'ч',
'ш', 'щ', 'ы', 'ь', 'э', 'ю', 'я', ' ']
alphabet1 =
['а', 'б', 'в', 'г', 'д', 'е', 'ж', 'з', 'и', 'й', 'к', 'л',
'м', 'н', 'о', 'п', 'р', 'с', 'т', 'у', 'ф', 'х', 'ц', 'ч',
'ш', 'щ', 'ы', 'ь', 'э', 'ю', 'я']
arr =
[0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,
19,20,21,22,23,24,24,26,27,28,29,30,31]
arr1 =
[0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,
19,20,21,22,23,24,24,26,27,28,29,30]
save = []
save1 = []
save2 = []

print("1. Letters with and without spaces\n2.
Bigram without crossing\n3. Bigram with
crossing")
ans = input()

def func (numb):
    for i in alphabet:
        letter = one.count(i)
        if letter == 0:
            P = 0
        else:
            P = -math.log2(letter/numb)
        H = letter/numb * P
        save.append(H)
        save1.append(i)
        save2.append("%.4f" % H)
    a = np.column_stack([save1,save2])
    b = sorted(a, key = lambda x: x[1])
    c = "\n".join(map(str,b))
    print(c)
    print("H1: ", (sum(save)))

def bigram(a, alp, numb):
    line = [one[k:k + 2] for k in range(0,
len(one), 2)]
    for i in alp:
        r = [alp[a] + i]
        bigram1 = line.count(r[0])
        if bigram1 == 0:
            P = 0
        else:
            P = -math.log2(bigram1/numb)
        H = bigram1/numb * P
        save1.append("%.3f" % H)
        save2.append(H)
    print(alp[a], " ".join(map(str, save1)))
    save1.clear()

def bigram2(a, alp, numb):
    line = [one[k:k+2] for k in
range(len(one))]
    for i in alp:
        r = [alp[a] + i]
        bigram1 = line.count(r[0])
        if bigram1 == 0:
            P = 0
        else:
            P = -math.log2(bigram1/numb)
        H = bigram1/numb * P
        save1.append("%.3f" % H)
        save2.append(H)
    print(alp[a], " ".join(map(str, save1)))
    save1.clear()

if ans == "1":
    # with spaces
    number = len(one)
    func(number)
    # without spaces
    print("\n")
    save.clear()
    one = one.replace(" ", "")
    number1 = len(one)
    func(number1)
elif ans == "2":
    # не перетинаються
    # with spaces
    number = len(one)/2
    print(" ", " ".join(alphabet))
    for item in arr:
        bigram(item, alphabet, number)
    print("H2 with spaces: ", (sum(save2)/2))
    save2.clear()
    print("\n")
    # without spaces
    one = one.replace(" ", "")
    number1 = len(one)/2
    print(" ", " ".join(alphabet1))
    for item in arr1:
        bigram(item, alphabet1, number1)
    print("H2 without spaces: ",
(sum(save2)/2))
elif ans == "3":
    # перетинаються
    # with spaces
    number = len(one)-1
    print(" ", " ".join(alphabet))
    for item in arr:
        bigram2(item, alphabet, number)
    print("H2: ", (sum(save2)/2))
    save2.clear()
    print("\n")
    # without spaces
    one = one.replace(" ", "")
    number1 = len(one)-1
    print(" ", " ".join(alphabet1))
    for item in arr1:
        bigram2(item, alphabet1, number1)
    print("H2: ", (sum(save2)/2))
else:
    print("Wrong input")
```

Значення H1 для тексту з пробілами:

['ф' '0.0193']
['ц' '0.0209']
['щ' '0.0227']
['э' '0.0308']
['ю' '0.0364']
['х' '0.0519']
['ш' '0.0522']
['й' '0.0579']
['ж' '0.0628']
['ч' '0.0832']
['з' '0.0871']
['б' '0.0872']
['ы' '0.0900']
['г' '0.0905']
['ь' '0.0919']
['я' '0.1091']
['п' '0.1303']
['у' '0.1369']
['д' '0.1393']
['м' '0.1428']
['к' '0.1528']
['в' '0.1571']
['р' '0.1746']
['с' '0.1937']
['л' '0.2004']
['т' '0.2171']
['и' '0.2238']
['н' '0.2264']
['а' '0.2546']
['е' '0.2624']
['о' '0.3227']
[' ' '0.4372']

H1: 4.366187862436842

Значення H1 для тексту без пробілів:

[' ' '0.0000']
['ф' '0.0225']
['ц' '0.0245']
['щ' '0.0265']
['э' '0.0360']
['ю' '0.0425']
['х' '0.0603']
['ш' '0.0607']
['й' '0.0672']
['ж' '0.0729']
['ч' '0.0961']
['з' '0.1006']
['б' '0.1007']
['ы' '0.1039']
['г' '0.1044']
['ь' '0.1060']
['я' '0.1255']
['п' '0.1495']
['у' '0.1568']
['д' '0.1596']
['м' '0.1634']
['к' '0.1746']
['в' '0.1795']
['р' '0.1990']
['с' '0.2200']
['л' '0.2274']
['т' '0.2458']
['и' '0.2530']
['н' '0.2559']
['а' '0.2865']
['е' '0.2948']
['о' '0.3586']

H1: 4.474445521025726

Значення H2 з пробілами без перетину:

H2: 3.978541281663438

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я		
а	0.000	0.008	0.022	0.004	0.017	0.011	0.011	0.028	0.001	0.006	0.039	0.059	0.023	0.022	0.000	0.009	0.022	0.027	0.030	0.002	0.001	0.010	0.001	0.010	0.008	0.003	0.000	0.000	0.000	0.010	0.014	0.101	
б	0.008	0.000	0.001	0.000	0.000	0.015	0.000	0.000	0.011	0.000	0.003	0.007	0.000	0.003	0.019	0.000	0.017	0.001	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.002	0.028	0.002	0.000	0.000	0.005	0.003	
в	0.032	0.000	0.001	0.000	0.004	0.033	0.000	0.005	0.024	0.000	0.002	0.005	0.002	0.012	0.039	0.003	0.008	0.022	0.004	0.006	0.000	0.000	0.000	0.001	0.005	0.000	0.016	0.001	0.000	0.000	0.003	0.040	
г	0.000	0.000	0.000	0.000	0.016	0.001	0.000	0.000	0.006	0.000	0.003	0.015	0.000	0.003	0.051	0.000	0.007	0.000	0.000	0.006	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	
д	0.034	0.000	0.008	0.000	0.000	0.041	0.014	0.000	0.021	0.000	0.003	0.008	0.001	0.015	0.025	0.001	0.008	0.002	0.001	0.016	0.000	0.001	0.003	0.000	0.001	0.000	0.006	0.005	0.002	0.000	0.004	0.012	
е	0.001	0.010	0.011	0.029	0.019	0.011	0.008	0.011	0.001	0.015	0.012	0.047	0.028	0.045	0.001	0.009	0.053	0.028	0.037	0.001	0.001	0.005	0.003	0.010	0.011	0.008	0.000	0.000	0.000	0.002	0.002	0.101	
ж	0.013	0.000	0.000	0.000	0.007	0.028	0.000	0.000	0.014	0.000	0.003	0.000	0.000	0.008	0.008	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.002	
з	0.042	0.001	0.007	0.004	0.008	0.003	0.002	0.000	0.004	0.000	0.002	0.002	0.002	0.016	0.005	0.000	0.002	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	
и	0.002	0.004	0.015	0.005	0.016	0.013	0.003	0.018	0.003	0.009	0.017	0.043	0.020	0.028	0.002	0.002	0.008	0.021	0.027	0.000	0.001	0.012	0.006	0.008	0.006	0.001	0.000	0.000	0.000	0.002	0.008	0.108	
й	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.004	0.000	0.000	0.000	0.004	0.003	0.000	0.000	0.000	0.001	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.049	
к	0.054	0.000	0.002	0.000	0.001	0.007	0.000	0.000	0.024	0.000	0.001	0.008	0.000	0.004	0.056	0.000	0.018	0.001	0.005	0.017	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.039	
л	0.042	0.001	0.000	0.002	0.001	0.035	0.003	0.003	0.046	0.000	0.004	0.005	0.000	0.003	0.044	0.000	0.000	0.017	0.001	0.011	0.000	0.000	0.000	0.002	0.000	0.000	0.008	0.008	0.027	0.000	0.013	0.073	
м	0.022	0.000	0.000	0.000	0.000	0.031	0.000	0.000	0.023	0.000	0.001	0.001	0.001	0.017	0.033	0.001	0.001	0.001	0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.053	
н	0.068	0.000	0.000	0.004	0.005	0.060	0.000	0.001	0.044	0.000	0.004	0.000	0.000	0.018	0.066	0.000	0.001	0.003	0.006	0.025	0.001	0.000	0.004	0.003	0.000	0.001	0.022	0.012	0.000	0.002	0.016	0.039	
о	0.000	0.026	0.041	0.039	0.031	0.016	0.019	0.011	0.009	0.029	0.020	0.041	0.040	0.054	0.002	0.011	0.039	0.041	0.047	0.001	0.008	0.006	0.001	0.020	0.010	0.002	0.000	0.000	0.001	0.003	0.000	0.122	
п	0.014	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.010	0.000	0.002	0.009	0.000	0.002	0.070	0.001	0.043	0.000	0.001	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.001	0.000	0.000	0.004	0.001	
р	0.044	0.002	0.004	0.001	0.003	0.040	0.004	0.001	0.036	0.000	0.003	0.001	0.004	0.010	0.046	0.001	0.005	0.013	0.009	0.032	0.000	0.001	0.000	0.002	0.001	0.000	0.014	0.006	0.001	0.002	0.000	0.009	
с	0.012	0.001	0.011	0.000	0.005	0.029	0.001	0.000	0.024	0.000	0.003	0.026	0.014	0.010	0.020	0.016	0.002	0.007	0.060	0.006	0.000	0.002	0.001	0.003	0.001	0.000	0.003	0.022	0.001	0.002	0.031	0.028	
т	0.042	0.001	0.015	0.000	0.001	0.035	0.000	0.000	0.027	0.000	0.007	0.002	0.002	0.009	0.088	0.002	0.022	0.009	0.003	0.015	0.000	0.000	0.000	0.001	0.000	0.000	0.013	0.040	0.001	0.002	0.006	0.044	
у	0.006	0.009	0.008	0.009	0.014	0.001	0.012	0.005	0.001	0.003	0.010	0.018	0.012	0.003	0.005	0.006	0.007	0.013	0.022	0.000	0.000	0.005	0.000	0.009	0.006	0.003	0.000	0.000	0.002	0.009	0.001	0.047	
ф	0.001	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.002	0.001	0.000	0.001	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
х	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.002	0.001	0.003	0.021	0.000	0.003	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	
ц	0.006	0.000	0.001	0.000	0.000	0.007	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	
ч	0.007	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.011	0.000	0.005	0.000	0.000	0.000	0.008	0.001	0.000	0.000	0.031	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	
ш	0.007	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.013	0.000	0.007	0.006	0.000	0.003	0.005	0.000	0.000	0.000	0.002	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	
щ	0.007	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.011	0.000	0.007	0.006	0.000	0.003	0.005	0.000	0.000	0.000	0.002	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	
ы	0.000	0.003	0.008	0.000	0.000	0.000	0.000	0.011	0.000	0.007	0.006	0.000	0.000	0.003	0.005	0.000	0.000	0.000	0.002	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	
ь	0.000	0.000	0.000	0.000	0.001	0.003	0.000	0.002	0.001	0.000	0.009	0.000	0.003	0.012	0.000	0.000	0.000	0.007	0.001	0.000	0.000	0.000	0.002	0.001	0.007	0.000	0.000	0.000	0.000	0.000	0.004	0.004	0.067
э	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.001	0.000	0.006	0.001	0.001	0.000	0.000	0.000	0.002	0.000	0.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
ю	0.000	0.003	0.000	0.000	0.004	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	
я	0.000	0.001	0.003	0.001	0.006	0.001	0.002	0.002	0.000	0.000	0.002	0.011	0.004	0.006	0.000	0.000	0.001	0.005	0.013	0.000	0.000	0.002	0.001	0.001	0.000	0.003	0.000	0.000	0.000	0.001	0.001	0.000	
	0.017	0.051	0.005	0.028	0.063	0.033	0.013	0.037	0.063	0.000	0.068	0.021	0.058	0.098	0.071	0.105	0.034	0.096	0.059	0.038	0.004	0.020	0.004	0.044	0.010	0.002	0.000	0.000	0.025	0.001	0.036	0.000	
H2 with spaces:	3.978541281663438																																

Значення H2 без пробілів та перетину:

H2: 4.1880373760474665

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я	
а	0.003	0.017	0.035	0.009	0.027	0.018	0.013	0.038	0.011	0.007	0.053	0.071	0.035	0.040	0.013	0.028	0.027	0.043	0.042	0.007	0.002	0.014	0.002	0.014	0.009	0.004	0.000	0.000	0.000	0.010	0.020	
б	0.010	0.000	0.001	0.000	0.000	0.017	0.000	0.000	0.013	0.000	0.004	0.009	0.001	0.004	0.023	0.000	0.020	0.002	0.000	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.032	0.002	0.002	0.000	0.007
в	0.038	0.004	0.003	0.003	0.009	0.038	0.001	0.007	0.028	0.000	0.008	0.007	0.005	0.017	0.047	0.008	0.010	0.031	0.010	0.008	0.000	0.001	0.001									

Значення H2 з пробілами та з перетином:

H2: 3.9801036688866813

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я		
а	0.008	0.008	0.021	0.004	0.017	0.011	0.010	0.029	0.001	0.006	0.039	0.059	0.023	0.022	0.008	0.009	0.021	0.027	0.030	0.002	0.001	0.010	0.001	0.010	0.007	0.003	0.008	0.008	0.008	0.010	0.010	0.002	
б	0.008	0.008	0.001	0.000	0.000	0.015	0.000	0.000	0.011	0.000	0.003	0.008	0.000	0.003	0.020	0.000	0.016	0.001	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.002	0.028	0.002	0.000	0.000	0.006	0.003	
в	0.032	0.008	0.001	0.000	0.004	0.033	0.000	0.005	0.023	0.000	0.002	0.005	0.002	0.012	0.038	0.003	0.008	0.022	0.003	0.006	0.000	0.000	0.000	0.000	0.001	0.005	0.000	0.017	0.002	0.000	0.000	0.003	0.041
г	0.008	0.008	0.008	0.000	0.016	0.002	0.000	0.000	0.006	0.000	0.003	0.016	0.000	0.003	0.051	0.000	0.007	0.000	0.000	0.007	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	
д	0.035	0.008	0.009	0.000	0.001	0.041	0.014	0.000	0.020	0.000	0.000	0.003	0.007	0.001	0.015	0.026	0.001	0.008	0.002	0.002	0.017	0.000	0.001	0.003	0.000	0.001	0.000	0.006	0.005	0.003	0.000	0.004	0.012
е	0.001	0.010	0.011	0.029	0.019	0.011	0.008	0.011	0.001	0.015	0.013	0.046	0.029	0.045	0.001	0.008	0.052	0.028	0.037	0.001	0.001	0.005	0.003	0.009	0.010	0.008	0.000	0.000	0.000	0.001	0.002	0.101	
ж	0.012	0.000	0.000	0.000	0.007	0.028	0.000	0.000	0.015	0.000	0.003	0.000	0.000	0.008	0.008	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.002	
з	0.042	0.001	0.007	0.004	0.008	0.004	0.001	0.000	0.004	0.000	0.002	0.002	0.002	0.002	0.005	0.000	0.003	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.003	0.017	
и	0.002	0.004	0.015	0.005	0.016	0.012	0.003	0.018	0.004	0.009	0.016	0.043	0.019	0.027	0.002	0.002	0.008	0.022	0.027	0.000	0.001	0.012	0.007	0.008	0.006	0.001	0.000	0.000	0.000	0.002	0.008	0.110	
й	0.008	0.008	0.008	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.001	0.004	0.000	0.000	0.000	0.004	0.003	0.000	0.000	0.000	0.001	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.048	
к	0.053	0.000	0.002	0.000	0.001	0.007	0.000	0.000	0.023	0.000	0.000	0.008	0.000	0.004	0.057	0.000	0.019	0.001	0.006	0.017	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.039	
л	0.042	0.001	0.000	0.002	0.001	0.035	0.003	0.003	0.045	0.000	0.004	0.005	0.000	0.003	0.043	0.000	0.000	0.017	0.001	0.011	0.000	0.000	0.000	0.001	0.000	0.000	0.008	0.028	0.000	0.009	0.014	0.072	
м	0.022	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.023	0.000	0.001	0.001	0.001	0.001	0.017	0.033	0.001	0.001	0.001	0.000	0.018	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.001	0.000	0.006	0.053	
н	0.067	0.000	0.000	0.004	0.005	0.061	0.000	0.001	0.043	0.000	0.005	0.000	0.000	0.017	0.065	0.000	0.001	0.003	0.006	0.025	0.001	0.000	0.004	0.003	0.000	0.001	0.023	0.012	0.000	0.003	0.016	0.039	
о	0.008	0.026	0.042	0.038	0.031	0.016	0.019	0.011	0.009	0.030	0.021	0.041	0.041	0.054	0.002	0.011	0.038	0.040	0.048	0.001	0.008	0.007	0.001	0.019	0.011	0.003	0.000	0.000	0.001	0.003	0.008	0.125	
п	0.015	0.008	0.008	0.000	0.000	0.000	0.023	0.000	0.000	0.009	0.000	0.002	0.008	0.000	0.002	0.070	0.001	0.043	0.000	0.001	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.001	0.000	0.004	0.001
р	0.046	0.002	0.004	0.001	0.003	0.039	0.003	0.001	0.036	0.000	0.004	0.001	0.004	0.010	0.045	0.001	0.005	0.013	0.010	0.032	0.000	0.001	0.000	0.001	0.001	0.000	0.013	0.006	0.001	0.002	0.009	0.009	
с	0.012	0.001	0.012	0.000	0.004	0.028	0.001	0.000	0.023	0.000	0.004	0.024	0.027	0.013	0.011	0.020	0.016	0.002	0.007	0.060	0.006	0.000	0.002	0.001	0.003	0.001	0.000	0.003	0.022	0.001	0.002	0.031	0.027
т	0.042	0.001	0.015	0.000	0.001	0.035	0.000	0.000	0.026	0.000	0.007	0.002	0.002	0.008	0.088	0.002	0.022	0.009	0.003	0.015	0.000	0.000	0.000	0.001	0.000	0.000	0.014	0.040	0.001	0.002	0.006	0.045	
у	0.006	0.009	0.008	0.009	0.014	0.001	0.012	0.004	0.000	0.002	0.010	0.018	0.012	0.003	0.005	0.006	0.007	0.013	0.021	0.000	0.000	0.004	0.000	0.009	0.006	0.003	0.000	0.000	0.002	0.009	0.001	0.046	
ф	0.001	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.002	0.001	0.000	0.001	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
х	0.009	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.002	0.000	0.003	0.021	0.000	0.003	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	
ц	0.005	0.000	0.001	0.000	0.000	0.007	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.003	
ч	0.018	0.000	0.000	0.000	0.000	0.028	0.000	0.000	0.013	0.000	0.005	0.000	0.000	0.008	0.001	0.000	0.000	0.000	0.032	0.006	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.003	0.000	0.000	0.000	0.001	
ш	0.007	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.011	0.000	0.007	0.007	0.001	0.003	0.005	0.000	0.000	0.000	0.002	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.001	
щ	0.007	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.011	0.000	0.007	0.007	0.001	0.003	0.005	0.000	0.000	0.000	0.002	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.001	
ы	0.000	0.003	0.007	0.001	0.001	0.008	0.001	0.001	0.000	0.013	0.001	0.018	0.011	0.002	0.000	0.002	0.003	0.004	0.007	0.000	0.000	0.007	0.000	0.002	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.038	
ь	0.000	0.000	0.000	0.000	0.001	0.003	0.000	0.001	0.000	0.000	0.009	0.000	0.003	0.012	0.000	0.000	0.000	0.007	0.001	0.000	0.000	0.000	0.002	0.001	0.007	0.000	0.000	0.000	0.000	0.004	0.004	0.067	
э	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.000	0.006	0.001	0.001	0.000	0.000	0.000	0.001	0.000	0.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
ю	0.000	0.003	0.000	0.000	0.004	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.003	0.000	0.000	0.001	0.000	0.003	0.000	0.005	0.000	0.000	0.000	0.001	0.000	0.021
я	0.008	0.001	0.004	0.001	0.007	0.001	0.002	0.002	0.000	0.000	0.002	0.011	0.004	0.006	0.000	0.000	0.001	0.005	0.013	0.000	0.000	0.002	0.001	0.001	0.000	0.003	0.000	0.000	0.000	0.001	0.001	0.001	0.081
	0.017	0.051	0.004	0.028	0.062	0.033	0.013	0.036	0.062	0.000	0.067	0.021	0.057	0.098	0.071	0.106	0.033	0.096	0.058	0.037	0.004	0.021	0.004	0.043	0.010	0.001	0.000	0.000	0.025	0.001	0.036	0.000	
H2:	3.9801036688866813																																

Значення H2 без пробілів та з перетином:

H2: 4.188308624728646

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я	
а	0.003	0.016	0.034	0.009	0.028	0.018	0.013	0.037	0.011	0.007	0.053	0.071	0.035	0.040	0.013	0.028	0.028	0.044	0.041	0.008	0.002	0.013	0.002	0.014	0.010	0.004	0.000	0.000	0.005	0.012	0.020	
б	0.009	0.000	0.001	0.000	0.000	0.017	0.000	0.000	0.013	0.000	0.004	0.009	0.001	0.003	0.023	0.000	0.019	0.002	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.002	0.032	0.002	0.000	0.000	0.007	
в	0.038	0.004	0.003	0.004	0.008	0.039	0.001	0.008	0.029	0.000	0.008	0.008	0.005	0.017	0.047	0.009	0.011	0.031	0.010	0.008	0.000	0.001	0.001	0.002	0.006	0.000	0.020	0.002	0.003	0.000	0.004	
г	0.010	0.001	0.001	0.000	0.019	0.002	0.000	0.001	0.007	0.000	0.004	0.019	0.001	0.005	0.059	0.002	0.009	0.002	0.001	0.009	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	
д	0.041	0.001	0.011	0.001	0.002	0.048	0.017	0.001	0.024	0.000	0.005	0.009	0.001	0.020	0.030	0.002	0.010	0.005	0.002	0.020	0.000	0.001	0.003	0.001	0.001	0.000	0.007	0.006	0.003	0.001	0.005	
е	0.003	0.020	0.026	0.037	0.029	0.016	0.011	0.019	0.010	0.018	0.023	0.023	0.056	0.040	0.063	0.014	0.027	0.063	0.044	0.049	0.008	0.001	0.009	0.003	0.016	0.010	0.000	0.000	0.003	0.002	0.007	
ж	0.014	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.018	0.000	0.004	0.000	0.000	0.010	0.009	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	
з	0.049	0.003	0.010	0.005	0.010	0.005	0.002	0.002	0.005	0.000	0.004	0.003	0.004	0.021	0.007	0.003	0.004	0.003	0.002	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.002	0.001	0.000	0.004	
и	0.005	0.013	0.033	0.009	0.028	0.019	0.005	0.025	0.017	0.011	0.027	0.051	0.029	0.046	0.016	0.024	0.013	0.041	0.039	0.008	0.002	0.017	0.008	0.014	0.008	0.001	0.000	0.000	0.004	0.003	0.015	
й	0.002	0.004	0.006	0.003	0.007	0.001	0.002	0.002	0.006	0.000	0.009	0.002	0.007	0.010	0.005	0.008	0.005	0.011	0.006	0.002	0.001	0.001	0.001	0.005	0.002	0.000	0.000	0.000	0.001	0.000	0.002	
к	0.062	0.003	0.006	0.002	0.005	0.010	0.002	0.002	0.030	0.000	0.005	0.010	0.005	0.009	0.068	0.006	0.023	0.007	0.010	0.021	0.000	0.001	0.001	0.003	0.000	0.000	0.000	0.000	0.002	0.000	0.002	
л	0.049	0.007	0.010	0.005	0.008	0.045	0.004	0.006	0.056	0.000	0.012	0.007	0.006	0.015	0.055	0.011	0.004	0.027	0.006	0.016	0.001	0.003	0.000	0.010	0.001	0.000	0.009	0.032	0.002	0.011	0.022	
м	0.027	0.004	0.007	0.003	0.006	0.038	0.001	0.002	0.031	0.000	0.007	0.003	0.005	0.026	0.041	0.011	0.003	0.011	0.004	0.022	0.000	0.001	0.000	0.005	0.001	0.000	0.018	0.001	0.002	0.000	0.010	
н	0.078	0.004	0.006	0.005	0.008	0.072	0.001	0.003	0.053	0.000	0.010	0.001	0.002	0.025	0.078	0.009	0.002	0.011	0.009	0.031	0.001	0.002	0.005	0.005	0.000	0.001	0.026	0.014	0.001	0.003	0.019	
о	0.002	0.038	0.062	0.047	0.044	0.023	0.024	0.018	0.019	0.035	0.033	0.050	0.055	0.076	0.019	0.033	0.048	0.061	0.065	0.008	0.010	0.011	0.002	0.030	0.014	0.003	0.000	0.000	0.007	0.003	0.014	
п	0.017	0.000	0.000	0.000	0.000	0.027	0.000	0.000	0.011	0.000	0.002	0.010	0.000	0.002	0.081	0.001	0.050	0.001	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.001	0.000	0.005	
р	0.053	0.003	0.006	0.002	0.006	0.046	0.004	0.001	0.043	0.000	0.006	0.002	0.005	0.012	0.053	0.002	0.007	0.016	0.012	0.037	0.000	0.002	0.000	0.002	0.002	0.000	0.015	0.007	0.002	0.002	0.011	
с	0.014	0.003	0.016	0.002	0.007	0.033	0.002	0.001	0.028	0.000	0.030	0.031	0.017	0.016	0.025	0.021	0.003	0.011	0.017	0.009	0.000	0.004	0.001	0.004	0.001	0.000	0.004	0.025	0.003	0.003	0.037	
т	0.050	0.003	0.022	0.002	0.005	0.042	0.001	0.002	0.033	0.000	0.011	0.004	0.006	0.015	0.084	0.009	0.027	0.017	0.008	0.019	0.001	0.001	0.004	0.001	0.000	0.004	0.016	0.046	0.002	0.003	0.009	
у	0.008	0.013	0.014	0.011	0.020	0.003	0.015	0.007	0.007	0.003	0.015	0.022	0.018	0.011	0.010	0.013	0.009	0.020	0.027	0.002	0.000	0.006	0.000	0.014	0.000	0.003	0.000	0.000	0.003	0.011	0.003	
ф	0.011	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.010	0.000	0.000	0.001	0.000	0.000	0.003	0.000	0.002	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
х	0.011	0.001	0.006	0.001	0.003	0.001	0.000	0.001	0.005	0.000	0.002	0.003	0.002	0.006	0.025	0.004	0.004	0.004	0.001	0.003	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.003	0.000	
ц	0.006	0.000	0.001	0.000	0.000	0.009	0.000	0.000	0.004	0.000	0.001	0.000	0.000	0.001	0.004	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	
ч	0.021	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.016	0.000	0.006	0.001	0.000	0.010	0.002	0.000	0.000	0.000	0.038	0.008	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.004	0.000	0.000	0.000	
ш	0.008	0.000	0.001	0.000	0.000	0.022	0.000	0.000	0.013	0.000	0.008	0.008	0.001	0.003	0.006	0.000	0.000	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	
щ	0.000	0.000	0.001	0.000	0.000	0.022	0.000	0.000	0.013	0.000	0.008	0.008	0.001	0.003	0.006	0.000	0.000	0.000	0.002	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	
ы	0.001	0.005	0.013	0.002	0.005	0.011	0.001	0.003	0.004	0.015	0.004	0.021	0.015	0.008	0.005	0.010	0.004	0.011	0.010	0.003	0.000	0.010	0.000	0.004	0.011	0.000	0.000	0.000	0.001	0.000	0.002	
ь	0.002	0.005	0.010	0.002	0.007	0.008	0.001	0.004	0.008	0.000	0.016	0.003	0.009	0.023	0.008	0.012	0.003	0.017	0.006	0.003	0.000	0.002	0.003	0.007	0.009	0.000	0.000	0.000	0.000	0.003	0.005	0.008
э	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.007	0.001	0.001	0.000	0.002	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
ю	0.000	0.004	0.003	0.001	0.006	0.001	0.002	0.001	0.002	0.000	0.004	0.001	0.002	0.003	0.002	0.003	0.004	0.005	0.005	0.001	0.000	0.001	0.000	0.006	0.000	0.006	0.000	0.000	0.001	0.001	0.001	
я	0.002	0.006	0.017	0.004	0.014	0.004	0.003	0.008	0.009	0.000	0.009	0.015	0.009	0.022	0.010	0.018	0.004	0.018	0.020	0.007	0.000	0.004	0.001	0.006	0.001	0.003	0.000	0.000	0.000	0.002	0.002	0.004
H2:	4.188398624728646																															

За допомогою програми CoolPinkProgram оцінили значення $H^{(10)}$, $H^{(20)}$, $H^{(30)}$. У кожному досліді було зроблено не менше ніж 50 експериментів.

Результат для $H^{(10)}$

Лабораторная работа №1

Произвольная часть текста:

вах_или_ч

Использованные буквы:

Порядок n-граммы:

5 символов

10 символов

15 символов

20 символов

25 символов

30 символов

35 символов

40 символов

45 символов

50 символов

Введенный символ:

Символ по счету:

Номер эксперимента:

51

Неравенство для энтропии:

$2.4289397293231 < H < 3.13669456263237$

Двоичная таблица угаданных символов:

00000100000000000000000000000000

00000000000001000000000000000000

00000000000000000001000000000000

00000000000100000000000000000000

00000000000000000001000000000000

Поле ввода символов:

Продолжить

Другой

Вероятности:

$q[1] = 0.42$

$q[2] = 0.1$

$q[3] = 0.08$

$q[4] = 0$

$q[5] = 0.06$

$q[6] = 0.02$

$q[7] = 0.02$

$q[8] = 0.02$

$q[9] = 0.02$

$q[10] = 0$

$q[11] = 0.02$

$q[12] = 0$

$q[13] = 0.06$

$q[14] = 0$

$q[15] = 0$

$q[16] = 0.02$

$q[17] = 0$

$q[18] = 0.02$

$q[19] = 0.04$

$q[20] = 0$

$q[21] = 0.02$

$q[22] = 0$

$q[23] = 0$

$q[24] = 0$

$q[25] = 0.02$

$q[26] = 0.02$

$q[27] = 0.04$

$q[28] = 0$

$q[29] = 0$

$q[30] = 0$

$q[31] = 0$

$q[32] = 0$

Строка состояния:

Результат для $H^{(20)}$

Лабораторная работа №1

Произвольная часть текста:

ругое_вы_заметили_ч

Использованные буквы:

Порядок n-граммы:

5 символов

10 символов

15 символов

20 символов

25 символов

30 символов

35 символов

40 символов

45 символов

50 символов

Введенный символ:

Символ по счету:

Номер эксперимента:

51

Неравенство для энтропии:

$2.17041016264485 < H < 2.87476470319971$

Двоичная таблица угаданных символов:

01000000000000000000000000000000

00000000000000010000000000000000

10000000000000000000000000000000

01000000000000000000000000000000

10000000000000000000000000000000

Поле ввода символов:

Продолжить

Другой

Вероятности:

$q[1] = 0.4$

$q[2] = 0.14$

$q[3] = 0.08$

$q[4] = 0.1$

$q[5] = 0.06$

$q[6] = 0.04$

$q[7] = 0.04$

$q[8] = 0$

$q[9] = 0$

$q[10] = 0$

$q[11] = 0$

$q[12] = 0$

$q[13] = 0$

$q[14] = 0$

$q[15] = 0.04$

$q[16] = 0$

$q[17] = 0$

$q[18] = 0.02$

$q[19] = 0$

$q[20] = 0$

$q[21] = 0.02$

$q[22] = 0$

$q[23] = 0$

$q[24] = 0$

$q[25] = 0.04$

$q[26] = 0$

$q[27] = 0$

$q[28] = 0$

$q[29] = 0.02$

$q[30] = 0$

$q[31] = 0$

$q[32] = 0$

Строка состояния:

Результат для $H^{(30)}$

Лабораторная работа №1



Произвольная часть текста:
 просто_стараясь_обратить_ваше_внимание_на_один_факт_а_именно_на_то_что_в_эт

Ипользованные буквы:

Порядок n-граммы:
 5 символов
 10 символов
 15 символов
 20 символов
 25 символов
30 символов
 35 символов
 40 символов
 45 символов
 50 символов

Введенный символ: **_ (пробел)**
 Символ по счету: **1**
 Номер эксперимента: **50**
 Поле ввода символов:

Продолжить
 Другой

Неравенство для энтропии:
 $1.43036504062232 < H < 2.17366068968819$
 Двоичная таблица угаданных символов:

10000000000000000000000000000000	▲
00000001000000000000000000000000	
00010000000000000000000000000000	
01000000000000000000000000000000	
10000000000000000000000000000000	▼

Вероятности:

$q[1] = 0,64$
$q[2] = 0,04$
$q[3] = 0,06$
$q[4] = 0,06$
$q[5] = 0$
$q[6] = 0,02$
$q[7] = 0,04$
$q[8] = 0,02$
$q[9] = 0$
$q[10] = 0$
$q[11] = 0,02$
$q[12] = 0$
$q[13] = 0$
$q[14] = 0,02$
$q[15] = 0,02$
$q[16] = 0$
$q[17] = 0$
$q[18] = 0$
$q[19] = 0$
$q[20] = 0,02$
$q[21] = 0$
$q[22] = 0,02$
$q[23] = 0$
$q[24] = 0$
$q[25] = 0,02$
$q[26] = 0$
$q[27] = 0$
$q[28] = 0$
$q[29] = 0$
$q[30] = 0$
$q[31] = 0$
$q[32] = 0$

 Строка состояния:
 Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

Результат надлишковості російської мови

Надлишковість джерела відкритого тексту (мови) дорівнює $R = 1 - \frac{H_x}{H_0}$

$$H_0 = \log_2 32 = 5$$

$$1 - (2.428/5) = 0.5144 > R^{(10)} > 1 - (3.136/5) = 0.3728$$

Аналогічні дії повели і для інших випадків.

$$0.566 > R^{(20)} > 0.4252$$

$$0.714 > R^{(30)} > 0.5654$$

Висновки: під час лабораторної роботи , ми закріпили свої знання стосовно етропії та надлишковості тексту. Вирахували значення для вхідного тексту для монограм та біграм. Удосконалили свої знання мови python для роботи з різними структурами даних та файлами, навчилися краще придумовувати алгоритми для роботи з вхідним текстом.